Creating Shared Instructional Products: An Alternative Approach to Improving Teaching

Anne K. Morris and James Hiebert

To solve two enduring problems in education—unacceptably large variation in learning opportunities for students across classrooms and little continuing improvement in the quality of instruction—the authors propose a system that centers on the creation of shared instructional products that guide classroom teaching. By examining systems outside and inside education that build useful knowledge products for improving the performance of their members, the authors induce three features that support a work culture for creating such products: All members of the system share the same problems for which the products offer solutions; improvements to existing products are usually small and are assessed with just enough data; and the products are jointly constructed and continuously improved with contributions from everyone in the system.

Keywords: educational reform; school/teacher effectiveness; teacher knowledge

In the February 15, 2009, edition of the *New York Times*, Nicholas Kristof pointed to our "greatest national shame": education. The reason for this unwanted distinction, according to Kristof, is that students' learning opportunities are much too uneven from school to school and classroom to classroom. Kristof's opinion has been voiced by others (Kozol, 1991), and observations of large variations in instructional quality have been widely reported (Pianta, Belsky, Houts, & Morrison, 2007; Raudenbush, 2009; Sanders & Rivers, 1996). Students who happen to receive ineffective instruction fall further and further behind those who receive good instruction. And some students receive very ineffective instruction (Hiebert et al., 2005; Sanders & Rivers, 1996; Stuhlman & Pianta, 2009). Even a few years of such instruction can doom students to a lifetime of fewer opportunities.

The large variation in students' classroom learning opportunities is made worse by the lack of improvement in the quality of school instruction over time. The core of teaching—the way in which a teacher and students interact about the content—has changed little in U.S. classrooms over the past century (Cuban, 1993; Fey, 1979; Hoetker & Ahlbrand, 1969). Of course, there have been changes at the margins, such as more colorful textbooks, and even the way the desks are arranged in some classrooms. But the nature of teaching—that which sets the real learning opportunities for students—has remained remarkably consistent over the years.

What can be done to solve these two problems: the large variation among classrooms and the lack of improvement over time? From our point of view, the problems are related. Making progress on the first provides the foundation for solving the second. In this article, we propose a way of representing these problems, and their solutions, that provides an alternative to most conventional approaches. To preview our proposal, we interpret the need to reduce variation in classroom instruction as not only a moral imperative but a scientific one as well. A significant obstacle to improving the quality of instruction on a large scale is the lack of consistency in the details of instruction from classroom to classroom. If instruction varies from classroom to classroom, there is no way of accumulating evidence about what works and what does not across different classroom settings. There is no way of developing a "science of improvement" (Kenney, 2008, p. 140) that yields the kind of knowledge needed to build improvements upon improvements over time.

By examining a range of systems both outside and inside education, we identified several critical features that allow professions to reduce the variation in performance across settings and to improve this performance incrementally but steadily in all settings over time. The fact that different systems and/or professions have developed similar features of their work environments to solve these problems suggests that these features are worth considering as educators search for more lasting solutions.

Creating Public, Changeable Knowledge Products

The professions and systems within professions that we examined succeed in reducing variation in performance across settings by creating shared, changeable knowledge products. The products are designed to help members of the profession solve the problems they share. The products are seen as containing the best current solution to the profession's problems. They both guide practice and provide a repository for the continuously accumulating knowledge about practice. As new knowledge is generated, the products are updated to guide even better solutions. Much of the work within the profession centers around improving these products and, thereby, the solutions to the most pressing problems of the profession. The products we identified range from

Educational Researcher, Vol. 40, No. 1, pp. 5–14
DOI: 10.3102/0013189X10393501

repair manuals in the Eureka system of the Xerox Corporation to medical treatment protocols in the health care profession to daily lesson plans for classroom instruction in education.

The concept of public products or artifacts holding shared knowledge can be found in analyses of how knowledge develops. Karl Popper (1972), for example, described three worlds of knowledge. The third world, of most interest here, consists of shared ideas treatable as public objects that can be jointly constructed, stored, and improved.[1] The products we have in mind have exactly this character. Applied to the teaching profession, for example, one could imagine treating lesson plans for daily instruction as products that could be shared and examined publicly, that could be improved over time through repeated trials in multiple classrooms (Hiebert, Gallimore, & Stigler, 2002; Snow, 2001). As more detailed knowledge is continually acquired about how to teach toward particular learning goals, the lesson plans are annotated and updated by embedding this knowledge into the plans.

System Features That Enable the Construction of Knowledge Products

The systems we examined share three features that appear to enable the development and refinement of jointly constructed knowledge products.[2] We summarize each of these features and then present several examples that show the features at work.

Enabling Feature 1: Shared Problems Across the System

The first enabling feature for developing jointly constructed knowledge products is sharing ownership of problems that the products help to solve. Unless all participants in the system are working to solve the same problems and unless solving them improves everyone's work lives, there is no compelling reason for them to work together to build problem-solving products.

Commitment to solving the same problems is displayed most clearly when participants are evaluated by their progress in solving these problems. When all participants are accountable, they are motivated even further to contribute to building products that help solve the problems. To gauge how far the United States is from creating school or education systems of the kind just described, imagine what it would be like if university researchers, curriculum developers, and classroom teachers all were evaluated by the same criterion: whether the joint products they created helped school students in their regions improve their learning.

Enabling Feature 2: Small Tests of Small Changes

The second feature addresses the research methods that are commonly used to test and revise shared products. As such, it smacks directly into the ongoing debate about appropriate research methods in education (e.g., the October 2008 issue of this journal, beginning with Bulterman-Bos's article; Jacob & White, 2002; Moss et al., 2009).

Our goal in the debate about research methods is quite modest. We simply wish to distinguish between large-sample randomized control trials that establish a treatment as more promising than its competitors and the empirical tinkering—the repeated small tests of small changes—required to refine and adapt a treatment to work effectively in multiple settings. We argue, with others, that a range of methods can and should be used to investigate different aspects of the knowledge-building enterprise (e.g., Lamberg & Middleton, 2009; Tharp & Gallimore, 1979, 1982). But we wish to highlight the value of small tests of small changes because (a) this set of approaches is most frequently associated with the testing and revising of the products we have in mind and (b) this kind of empirical tinkering often is ignored or dismissed in descriptions of *scientific* approaches to knowledge building.

Perhaps small tests of small changes are used so often in the systems we examined because they are well suited to a "science of improvement" (Kenney, 2008, p. 140). For example, consider a lesson plan for classroom instruction that includes a prescription of actions the teacher should perform, the rationales for why particular actions are recommended, and questions regarding unresolved issues that still must be tested. Improvements in the lesson are most likely to occur through making small changes in the lesson and checking the effects. Does this task yield better learning opportunities than that one? Does this question generate more productive thinking than that one? As the lesson plan is tested in multiple classrooms, it is most efficient to conduct small tests—collecting just enough data to evaluate the effects of the change—and then accumulate the results over multiple trials.

Replication of small trials is not a foreign concept in scientific research. In fact, as described by Tukey (1969), Sir Ronald Fisher, the father of modern significance testing, saw the repetition of results, rather than the single result from a large study, as the real measure of scientific work. Consistent with this view, Karl Popper (1944/1985) outlined a scientific approach for *improving* interventions in the social sciences that involved making and testing small adjustments. This "piecemeal tinkering" approach (p. 304), said Popper, yields small mistakes, and it is easier to learn from small mistakes than large ones. Repeated small tests and learning from small mistakes provides the surest way to continually improve an intervention. Nathaniel Gage (1989) built directly on Popper's views when envisioning new (scientific) methods for studying teaching, a kind of empirical tinkering, which would break down the old walls between research and practice and enable useful knowledge for teaching to emerge.

Enabling Feature 3: Multiple Sources of Innovation

One reason that *jointly* created products can effectively help to solve a profession's problems is that the knowledge used to build the products is harvested from participants across the system, participants who likely possess different kinds of knowledge. Taking advantage of different kinds of knowledge and different kinds of expertise results in products that are more useful and of higher quality than products created by individuals working alone (Argyris & Schön, 1996). In addition, products that are jointly constructed are owned by all the participants, which, in turn, results in increased use of the products and increased commitment to improve them over time (Bobrow & Whalen, 2002; Douthwaite, 2002; Lewin & Grabbe, 1945).

Taking advantage of knowledge distributed across the system requires that members of the system respect each other's ideas. What one work group contributes must be viewed by the other work groups as essential for creating the best product possible.

Examples of Systems That Build Public, Changeable Knowledge Products

The usefulness of public, changeable knowledge products for reducing variation across settings and improving performance over time, and the critical role played by the three features we have identified all come to life when seen at work across a range of professions. We highlight two examples in this section, one outside education and one inside education. Then we pull together several major points regarding the nature of instructional products and how they can help solve the two big problems: reducing variation and improving performance over time. We conclude our analysis with a third example that helps sharpen the ideas we are proposing.

The Quality Movement in Health Care

We begin the description of our first example—continuously improving clinical medical care—by noting that education is different in many ways from other professions, including medicine. Often, in professions outside of formal education, the goals can be stated more clearly, the variables affecting the outcomes can be identified with greater certainty, and the outcomes can be measured with simpler assessments. This means that the processes used to build knowledge in these systems cannot simply be imported into education with an expectation of equal success. But an analysis of other systems can help educators see their own system more clearly and ask whether more productive knowledge-building processes are possible. That some systems outside education share our assumption that reduced variation and improved performance come from humans collaborating to build increasingly effective knowledge products (Argyris & Schön, 1996) makes them useful mirrors to reflect on the education system in the United States.

The health care example begins in the late 1980s, when Dr. Donald Berwick (recently appointed by President Obama as administrator of the Centers for Medicare and Medicaid Services) and a few other health care leaders became increasingly concerned about the quality of health care delivery in the United States. Noting too many preventable errors and more variability in treatment than research indicated, these leaders began searching for ways to improve the quality of care (Kenney, 2008). They studied the processes used by industry to improve the quality of manufactured products and saw their own products—treatment protocols—in a new light.

For example, Berwick and others noted that, contrary to popular perception, relatively little health care performance is actually measured. If treatment protocols are not described in detail and if the outcomes are not measured, there is no way to know whether they are working. George Halvorson, chief executive officer of Kaiser Permanente, noted that rather than being based on scientific evidence, the majority of health care in the United States is based on "local practice, habit, intuition, or dependent upon the training of a particular clinician" (Kenney, 2008, p. 242).

The key to improvement, decided Berwick and others, was standardizing treatment protocols that aimed to accomplish the same goal and then measuring the outcomes using the same assessments. This would allow recurring test-and-revise cycles and produce increasingly effective treatments. Such cycles are impossible when, as in one hospital, 50 variations were recorded in the process of inserting IVs (Kenney, 2008, p. 116). With this variation, there was no way of finding the sources of the common IV infections. In contrast, the eye-catching success of the Intermountain Hospital Network in improving health care across the board can be attributed, in large part, to the standardization, testing, and continual refinement of treatment protocols (Leonhardt, 2009). Dr. Brent James, director of the Intermountain Hospital Network, tried to convince the doctors in the network that "it's more important that you do it the same way than what you think is the right way" (Leonhardt, 2009, p. 33). It is more important, that is, for the ultimate improvement of treatment throughout the network because it allows study and improvement.

Frequently, doctors who are asked to become part of this kind of improvement system resist because they feel it turns medicine into a mechanical system and turns them into little more than robots. Patients, say the doctors, are too variable for this system to work. Dr. Lucian Leape, a leader of health care quality improvement initiatives, responds, "It's not the patient variations we're talking about. It's *our* variations that are the problem" (Kenney, 2008, p. 70).

The logic behind Dr. Leape's observation is that variability is impossible to study if the variability among patients is used as a reason to treat each patient differently according to the doctor's intuitive judgment. To be studied, variability must arise as the differential effect of standardized treatments. As different groups emerge, defined by the similarity of their responses to a treatment, the reasons for the variability between groups can be studied systematically and the treatment can then be adapted to meet each group's needs.

A linchpin of the activity for improvement in health care quality is the Institute for Healthcare Improvement (IHI), begun by Donald Berwick and colleagues in 1991. IHI runs numerous workshops and courses for those interested in improvement processes and also hosts a website that shares what the institute is learning about improving health care. One set of documents addresses measuring changes in treatment protocols to determine whether they are improvements or just changes. The documents propose a distinction between measuring for improvement and measuring for research (IHI, 2007). Measurements for research are critical for improvement, say the documents, but play a special role. They involve large-scale randomized control trials that are costly and take time to get results. In contrast, argues the IHI, measurements for learning and improvement consist of many sequential small-scale tests of a change (e.g., running tests of a change on the next 10 patients who enter a hospital). Measuring for improvement involves collecting just enough data to learn from the test, and the new knowledge is used to plan the next test.

Dr. Uma Kotagal, a longtime advocate for continuously improving health care who practices at Cincinnati Children's Hospital, noted the tension between methods of research commonly accepted in academic circles and those better tuned to improving clinical practice. "We learned a lot about how difficult it was to do this [improving quality] in healthcare and how difficult it was to do in an academic organization where research was the currency and fixing systems was viewed as somewhat sloppy and fuzzy," Kotagal said. "We knew at that point in time that we had to create a respect for the *science of improvement* [italics added] for it to survive in the organization" (Kenney, 2008, p. 140).

The just-enough-data approach has a number of advantages over using one large test to fully answer a complex research question (Langley, Nolan, Nolan, Norman, & Provost, 1996). Multiple cycles of small tests of small changes in treatment protocols reduce risks and costs as changes move from hunches to hypotheses to revisions. Small-scale tests can accelerate the rate of improvement because they allow a change to be empirically tested and quickly refined. Small tests of small changes also are less likely to disrupt the existing structure of the organization and can decrease organization members' resistance to change (e.g., Nonaka & Takeuchi, 1995). Changes are more likely to be accepted by participants because not only are they relatively small but they come with the assurance that they will be modified and tested again as needed. Over time, of course, small changes can combine to produce large improvements in a system.

A final observation about the quality movement in the health care arena is the democratization of input. Innovations, including error-prevention strategies, are invited and welcomed from throughout the system. All members, from technicians to nurses to doctors, are included in the process of studying and improving treatment protocols. Kenney (2008) noted that a key to the world-renowned success in improving health care delivery and eliminating almost all routine errors at the medical centers in Jönköping County, Sweden, was a "leveling of the hierarchy: a cultural shift that embraces the value of all contributors no matter what their position" (p. 227). An American visitor noted, with astonishment, "We walked into a meeting in Jönköping and we could not tell who was the doctor, who was the nurse, who was the tech, who was the medical assistant" (Kenney, 2008, p. 227). In the world of health care, that is a substantial cultural shift.

It should be apparent that the development of treatment protocols, the products that guide the reduction in variation and the improvement in treatment over time, is made possible by the three features described earlier. First, members of the health care system share the basic problem of improving the outcomes of clinical care. Because all members see themselves as, ultimately, contributing to the care of patients, improving care means they all are solving their primary problem.

Second, the information needed to improve treatments is gathered through small tests that yield just enough data. In health care, this science of improvement requires standardizing treatment protocols and using common assessment measures so that the sources of errors and the best protocols for treatments can be identified over multiple trials.

Finally, the system brings together the contributions of individuals working at different places in the system and possessing different kinds of knowledge. This democratization of input allows knowledge from throughout the system to be fed into revised treatment protocols.

Lesson Study in Japanese Schools

Our second example of a system that builds public, changeable knowledge products comes from within education. The system of lesson study in Japan has received considerable attention over the past few years (Lewis, 2002; Shimahara & Sakai, 1995; Wang-Iverson & Yoshida, 2005) and is quite well known, so we will describe it only briefly. It is of special interest here because, over the past 60 years, it has grown into a countrywide system and is seen by many as largely responsible for the high-quality teaching in Grade 1–8 Japanese classrooms (Lewis & Tsuchida, 1997; Shimahara & Sakai, 1995; Stigler & Hiebert, 1999).

The lesson study process usually involves three to six teachers working together for several months, often with an outside expert, to study and improve a daily lesson. It includes a cycle of (a) specifying the learning goals for the lesson; (b) searching the literature, consulting experts, reviewing past lesson plans, and discussing alternative instructional approaches for helping students achieve the learning goals; (c) using these resources to design a lesson by creating a detailed lesson plan; (d) teaching the lesson in one classroom while other teachers (plus visitors, on occasion) observe the lesson, noting whether it facilitates the predicted learning by the students; (e) debriefing about the success of the lesson, often hearing from all observers; and (f) revising the lesson based on how it helped (or did not help) students to achieve the learning goal or goals, and usually teaching it a second time and revising again.

The lesson study cycle includes a sequence of carefully choreographed events designed to increase these teachers' knowledge and the profession's knowledge of how to help students reach these particular learning goals. The process illustrates, in a rather inspired fashion, the features of product-building systems identified earlier.

The lesson plans constitute the public, changeable knowledge products that guide actions toward helping students achieve the learning goals. The plans suggest how to both teach the lesson *and* present the shared knowledge about how best to do so. Revised lessons update this knowledge, often by including the rationale for changes. Representing the knowledge for teaching toward learning goals in lesson plans places the knowledge at teachers' fingertips when needed, immediately accessible with minimal translation. Knowledge of why particular instructional decisions were made rather than just prescriptions for teaching helps teachers understand the decisions and, in turn, adapt the lessons to their local contexts. Being concrete and public, the lesson plans invite testing and further revising by other teachers. Consequently, the products and the knowledge they contain are considered changeable: tentative, always open to improvement.

The joint construction of the lesson plans is enabled by the fact that how to help students achieve the lesson's learning goals constitutes the shared problem for the group. The lesson's learning goals often are defined with larger learning goals in mind, goals often proposed by the National Ministry of Education. Lessons sometimes are designed specifically to show how these larger goals can be "brought to life in the classroom" (Lewis & Tsuchida, 1997). That is, the lesson study process helps teachers gain a shared understanding of these learning goals.

Because of national-level learning goals and a de facto national curriculum, the problem of helping students achieve these learning goals is shared not just by the three to six teachers in the local group but by all teachers in the country who teach the same subject and grade level. This means that what this group learns about the lesson goals and about teaching toward them will be sought out by many teachers across the country. Shared learning goals also means that these jointly developed lessons are treated as highly portable. Teachers who hold the same learning goals for students seek out well-crafted lessons developed by others.

As in the health care example, the lesson study process depends on small tests of small changes. The success of lessons often hinges on small changes—asking a question in just the right way at the right time or posing a particular task at a particular point in the lesson. Lesson study focuses on lessons at this level of detail and observes their implementation by monitoring students' responses at key moments in the lesson. That is, members agree on the kind of data that will most directly address the effectiveness of the lesson. Data gathered from one classroom and one trial often are enough to suggest a slight change in the lesson. The second trial often is designed to test this small change. Replicated over time by many study groups in many classrooms, knowledge of the effects of these small changes gradually accumulates and moves practice toward the versions of the lessons that most effectively help students achieve the learning goals. As teachers gravitate toward similar lessons, the variation in instruction across classrooms is reduced.

Finally, lesson study solicits innovations from educators throughout the system: from many teachers, university researchers, and national educators. The system is designed to harvest the best ideas from all groups and steadily feed these into lesson plans that are tested, revised, and gradually come into wider use.

In summary, lesson study in Japan is a case of building knowledge-rich products that reduce variation and improve teaching across the system. The construction of these special annotated lesson plans is enabled by the same supporting features we described earlier.

The Nature of Instructional Products

With the examples of improving health care by studying treatment protocols and improving teaching by studying lesson plans, we can now summarize several key points and extend our argument by making several additional observations. We begin by defining what we mean by *instructional products*. Because we set out to address the dual problem of variation in instructional practice and the lack of improvement in such practice over time, we focus on the instructional products for teaching that we believe most directly affect classroom practice: (a) specially annotated instructional (lesson) plans designed with specific learning goals in mind that describe all aspects of classroom instruction believed to affect students' opportunity to achieve the learning goals and that contain all the information needed to implement the plan as described and (b) assessments that measure whether students are achieving the learning goals. The kind of annotated lesson plans we have in mind are rich enough and complete enough to constitute a fully developed curriculum. This definition is consistent with the products of the lesson study process just described. Other supporting instructional products, outside the scope of this article, can be developed through a similar process, products such as special diagnostic tests for students and school-level tools to improve alignment among school variables (e.g., Strategic Education Research Partnership, n.d., "Internal Coherence Partnership and Protocol").

Key Characteristics of Useful Instructional Products

The purpose of the instructional products we describe is to help teachers help students achieve specified learning goals. Consequently, a first characteristic of useful instructional products (we will call them *annotated lesson plans* and *assessments*) is that they are created around particular learning goals. The products are useful for teachers if and only if teachers are trying to help students achieve these learning goals. Another way of saying this is that learning goals for students are the *shared problems* of instructional product systems. (Recall Enabling Feature 1 of systems that create useful products.) All members of the system view their primary task—helping students achieve the agreed-upon learning goals—as a shared problem.

A second characteristic of useful instructional products is that they are at a grain size that can guide classroom instruction. Because we are proposing instructional products that can reduce variation in instruction across classrooms and improve instruction over time, the products must be detailed enough to directly affect practice.

A third characteristic of useful instructional products is that they are testable and improvable. They must be tested against students' achievement of the learning goals. Using small tests of small changes, information can be gathered for revising the lesson plan and testing it again. Repeatedly testing lesson plans using empirical data on students' learning provides a mechanism for vetting their quality.

A fourth characteristic of useful instructional products is that they are accessible to teachers when needed. In the current U.S. system, most research-based information is difficult for teachers to access on demand because it often is stored on library shelves. Well-specified learning goals provide an answer to this challenge by indexing annotated lesson plans according to learning goals. The Internet provides many attractive possibilities for storing, searching, and accessing annotated lesson plans and assessments when needed.

The four characteristics we have identified are apparent in the examples of treatment protocols produced through health care improvement and annotated lesson plans produced through lesson study. The characteristics work together to reduce variation in instruction from classroom to classroom and improve students' learning over time. Shared learning goals mean that teachers will be working toward the same outcomes, thereby reducing variation due to differences in intended goals as well as providing a target against which changes in practice can be measured. Ensuring access to annotated lesson plans when needed, plans that guide instruction at a detailed level, increases the chances that teachers will be implementing the same instruction. Finally, testing the effectiveness of the implemented plans provides the evidence needed to improve the plans over time.

These four characteristics also mean that instructional materials such as state teaching and learning standards, scope and sequence charts, and textbooks with little lesson-level detail do not fit our definition of instructional products. They are not always indexed to specific learning goals, the grain size of the content does not guide daily instruction, and their influence on students' achievement of specific learning goals cannot be measured directly.

Using Instructional Products as Intended

Variation in instruction among classrooms is introduced not only by teachers teaching toward different goals using different plans but also by teachers implementing the same plans in different ways. Implementing the same plan in different ways might be

due to differences in the expertise of the teacher or differences in the context that prompt teachers to change the plans. How does the system we propose handle these variations?

Variations due to differences in expertise. Annotated lesson plans contain knowledge of two kinds: what to do and why/how to do it that way. What to do offers prescriptions that teachers can implement; why/how to do it that way provides a rationale or local theory for why the prescription might work, along with information that teachers will likely need to implement the plan as described.

Educators have recognized that teachers, as well as students, can and must learn from curricula if they are to teach with increasing effectiveness (Ball & Cohen, 1996; Davis & Krajcik, 2005; Remillard & Bryans, 2004). Annotated lesson plans include information for teachers that is designed to help novices implement the lessons as effectively as experts. Just as the prescriptions for what to do are improved over time as evidence is collected regarding the effects of the lesson on students' achievement of the learning goals, information for teachers for implementing the plan is supplemented and improved over time as novices implement the lesson and evidence is gathered regarding aspects of the lesson that were and were not understood well by the teachers and were and were not implemented as intended. Variations will continue to exist because the information cannot remove all differences in the way teachers will implement lessons even if they wish to implement them in similar ways, but these unintentional variations can be reduced over time.

Variations due to contextual differences. Variations in how lessons are implemented using the same annotated lesson plans might be due to differences in classroom variables, such as differences in the way students respond to the instructional activities. It is likely that some variations in implementation are appropriate to most effectively help these students achieve the learning goals. How does the system we propose deal with variation due to these factors?

First, it is important to remember the earlier health care example and Dr. Lucian Leape's observation (Kenney, 2008, p. 70), which we rephrase here to apply to teaching: It is not the variation among students that is the problem; it is the idiosyncratic variation among the instructional treatments that prevents accumulating knowledge about better instruction. Consequently, it makes sense to study standardized instructional treatments across variable settings, examining the data to learn about the effects of the contextual variables in the settings. Understanding the effects of the contextual variables then enables adapting the treatment to fit the context. We conjecture that patterns will emerge that suggest several variations of annotated lesson plans keyed to local conditions.

That instructional products will likely be adapted for local settings means that the rationales for these decisions are an important part of the knowledge contained in the products. The rationales for the prescriptions, or the small, local theories about why particular activities work well in this setting (and better than other activities that have been tried), provide information teachers can use to fine-tune the products even further for their contexts without introducing "lethal mutations" (Brown, 1992, 1994). These theories-of-action, or "practical arguments," provide an arena within which teachers can reason about their practice (Fenstermacher, 1988).

Two additional observations about instructional products. We wish to point out that developing products that serve this dual purpose—of prescribing instruction that works well in actual classrooms, and proposing local theories that explain the instructional successes and guide adaptations to fit other contexts and improve over time—is characteristic of design research (Cobb, Confrey, DiSessa, Lehrer, & Schauble, 2003; Collins, Joseph, & Bielaczyc, 2004; Kelly, 2003; Lamberg & Middleton, 2009). The similarities reinforce our suspicion that the product-building systems we are describing are consistent with, rather than at odds with, other emerging efforts to describe processes that create useful knowledge for classroom teaching.

A final observation about instructional products addresses the question of whether they support *good* teaching. We have not defined good teaching; indeed, what counts as good teaching is a contested issue in education (Fenstermacher & Richardson, 2005). Even if the knowledge-building system works as we envision, educators are likely to disagree about whether instruction is improving. In our view, good teaching can be defined only in relation to specific learning goals. In the system we propose, whether one judges teaching to be good and/or improving depends entirely on whether one values the learning goals toward which teaching is directed. If the learning goals are valued and students are achieving them more effectively, then teaching is improving. The system we propose is neutral with respect to learning goals; it is designed to improve the effectiveness of teaching with respect to any specified learning goals.

One More Example

We use our final example, the Strategic Education Research Partnership (SERP) (Donovan, Wigdor, & Snow, 2003), to clarify the ideas we are proposing. Because SERP is consistent in many ways with the ideas we have proposed but is different in some, we can sharpen our proposal by describing these similarities and differences.

Created in 2003 by the National Academies with financial support from a wide range of foundations, the SERP Institute aims to develop programs of research and development, staffed by teams of researchers and teachers, that address problems of genuine importance to schools. Here, we spotlight Word Generation, one of the first programs of SERP, as we compare SERP's goals, products, and features with those we have proposed.

Word Generation is a middle school academic language program designed to help students learn the vocabulary they will need to succeed in academic courses in high school and beyond (Snow, 2010; White & Kim, 2009). The program originated in a collaboration among researchers at Harvard University, SERP, and the Boston Public Schools. Word Generation has been through several rounds of pilot testing and revision and is now available to schools around the country. The program consists of 24 weeks of 15-minute modules that are taught by different subject-matter teachers each day of the week.

To compare the development of Word Generation and other SERP programs with the processes we have described, we first consider the features of the R & D system that build the products. We then ask whether and how the development of

SERP programs addresses the two major problems of too much instructional variation among classrooms and no meaningful improvement over time.

Features of the SERP R & D System

Feature 1: Shared problems across the system. Word Generation, along with other SERP programs, places considerable emphasis on identifying problems that are shared by all members of the R & D teams. This is accomplished by asking schools and teachers to identify problems they want to solve, asking researchers to adopt these problems, and then creating R & D teams that include members from all stakeholder groups (Strategic Education Research Partnership, n.d., "The SERP Field Site Structure"). In the case of Word Generation, teachers in Boston's secondary schools told SERP researchers that their students struggled because they stumbled on academic or technical vocabulary (Viadero, 2009).

It is instructive to review several issues that SERP has experienced as teachers and researchers pursue shared problems, because they are issues that are likely to arise with all shared problems in education. First, the Word Generation program illustrates the crosscutting general nature of many SERP R & D problems (Strategic Education Research Partnership, n.d., "The SERP Field Site Structure"). These problems are not the daily, instructional problems of individual teachers but rather are more general problems that cut across grade levels and subject areas. This allows the problems to be shared by more teachers and the interventions to be independent of local curricula. But this also means that the problems are not especially pressing for many teachers. The less-than-pressing nature of the problems prompted SERP researchers to require few if any changes in the teachers' instructional routines to implement the interventions (e.g., Strategic Education Research Partnership, n.d., "Algebra I"; n.d., "An Introduction to Word Generation"). The low demand on teachers' time and energy increases the chances the teachers will implement the intervention. But the low demand likely decreases the chances that teachers will take ownership of the continuous testing and refining of the intervention and make it part of their daily work lives. Although the instructional products we defined are necessarily a more central part of the affected teachers' work lives, the SERP experiences make clear that fitting the implementation and testing of instructional products into teachers' ongoing daily routines is a challenge for the approach we are proposing.

The Word Generation case reveals a different challenge for researchers. The problems selected for R & D are not the kind usually tackled by researchers. The problems are too close to the details of classroom instruction, and revising the interventions requires ongoing empirical tinkering using feedback from teachers (Strategic Education Research Partnership, n.d., "The SERP Field Site Structure"). To turn the problems into research data that offer chances for professional advancement (i.e., publication), researchers must wait for the experimental comparisons that come nearer the end of SERP's R & D cycle. As we suggested earlier, and Cochran-Smith and Lytle (1999) noted directly, academic cultures will need to change considerably to make researchers interested in the small tests of small changes involved in implementing and improving instructional products.

Feature 2: Small tests of small changes. As we have just described, the SERP cycle of R & D appears to use small tests of small changes in the early phases, while interventions and instruments are being field-tested and revised. Perhaps to give the products scientific credibility and/or to entice researchers to participate in the process, the later phases involve conducting controlled comparisons of the products (e.g., Snow, Lawrence, & White, 2009; White & Kim, 2009). These kinds of comparisons are unlikely to provide the kind of detailed feedback needed to improve the products, but they do yield publishable reports. Although we do not deny the value of controlled comparisons, we have argued that carefully planned empirical tinkering—small tests of small changes replicated over multiple settings—can have as much scientific credibility as controlled experiments do. In addition, small tests of small changes can provide continuing information for constantly updating the instructional products.

Feature 3: Multiple sources of innovation. The SERP R & D process is structured to "flatten the culture," to respect the ideas of all members of the system, to take advantage of the different knowledge held by different kinds of educators. The design team, for example, includes researchers, designers, and teachers who collaboratively develop instructional interventions, assessments, and data management tools to address the selected educational problem (Strategic Education Research Partnership, n.d., "The SERP Field Site Design Team"). Word Generation illustrates the way in which research can inform the pedagogical details of instructional practice and blend research findings with key instructional moves. In general, the features of the SERP design work demonstrate explicit efforts, consistent with our proposal, to blend the types of knowledge that are held by different educators and usually kept separate. Also consistent with our proposal, outlined in a later section, is the fact that different groups of educators can play different roles but still integrate their knowledge into jointly constructed products.

Solving the Problems of Too Much Variation and Not Enough Improvement

Reducing variation in classroom instruction intended to achieve the same learning goals is not an explicit goal of SERP programs. In fact, Word Generation is designed to allow flexible use by teachers (Strategic Education Research Partnership, n.d., "Word Generation: Teacher Guide"), and it appears that teachers implement it in different ways (White & Kim, 2009). But given the creation of portable products and the gathering of feedback from teachers in different settings, it appears that the products, and their implementations, could eventually lead to smaller instructional variations across classrooms.

Continuous improvement of instructional products over time is an explicit goal of the SERP R & D process (Donovan et al., 2003). Consistent with our proposal, SERP sees a long-term steady improvement process as more viable than the quick fixes to which many Americans are addicted. Continuous improvement requires that the products remain open to change. As the Word Generation instructional modules were released, a website invited teachers to give feedback so that researchers could improve them over time (Word Generation, n.d., "Exchange Ideas"). "We don't see this as an end product that needs to be sold and,

therefore, defended," said Suzanne Donovan, executive director of the SERP Institute. "As soon as you commercialize something, you have to defend it. And we don't want to say, 'This is it'" (Viadero, 2009, p. 1).

Different Roles Played by Different Educators

A final issue we address was raised by examining the SERP R & D process: the roles played by different members of the educational system. Currently, educational professionals with different kinds of knowledge create different products (research articles, textbooks, lesson plans). Students rarely benefit from the scope of knowledge stored separately in different forms and places. Furthermore, few venues exist that enable the integrating or blending of knowledge in forms that teachers actually use and students actually encounter. There is simply no way that the most effective instructional products can develop in a system that segregates the products created by people with different kinds of knowledge.

Different products are a direct consequence of the fact that different groups in the educational system are trying to solve different problems. Due to the differences in problems addressed and methods used, the kinds of knowledge produced by different groups, say researchers and teachers, are often seen as qualitatively different (Clandinin & Connelly, 1991; Cochran-Smith & Lytle, 1990, 1993; Eisner, 1995; Kennedy, 1999). At the least, the products created by these groups are separate, with little, if any, overlap or integration.

What would it look like if different groups used their special kinds of knowledge to solve different aspects of the same problems? The SERP example we reviewed earlier hints at how groups with different knowledge could blend their efforts to jointly construct useful products. To be explicit, we provide the following descriptions simply to note that such distributions of effort are possible, although numerous other configurations can be imagined:

1. Clarify and interpret learning goals.
   - Based on students' responses, teachers might further unpack the components of the lesson learning goals (proposed by curriculum writers) to identify the concepts and skills students need to achieve the goals (Morris, Hiebert, & Spitzer, 2009).
   - Researchers might examine the connections among the learning goals across lessons and create coherent sequences or progressions of learning goals (Corcoran, Mosher, & Rogat, 2009).
2. Develop curricula and lesson plans.
   - Curriculum writers, with the assistance of researchers, might develop first drafts of lesson-level curricula that incorporate relevant research findings.
   - Teachers and researchers might develop testable conjectures for improving curricula based on repeated tests across multiple classrooms (El Barrio–Hunter College PDS Partnership Writing Collective, 2009; Hiebert et al., 2002).
3. Develop local instructional theories.
   - Teachers might test local theories by implementing and then refining and modifying them, identifying the effects of local contextual conditions (Gravemeijer & van Eerde, 2009; Lamberg & Middleton, 2009).
   - Researchers might create local theories by taking into account the effects of instructional interventions across individual classrooms (Gravemeijer & van Eerde, 2009; Stigler & Thompson, 2009).
4. Design and apply assessments to gather relevant student data.
   - Teachers might use informal assessments to provide immediate feedback on the effectiveness of particular instructional activities (Wiliam, 2009).
   - Researchers might create common assessments to test the shared products used across classrooms to enable sharing data from replications and assessing improvements under different contextual conditions (Gage, 1989).

It is clear that teachers, researchers, and curriculum developers will bring different kinds of knowledge to these tasks. Although the boundaries are slippery, it seems clear that researchers or outside educators will need to fill in subject matter expertise and research design expertise, whereas teachers will need to provide classroom delivery (pedagogical) expertise as well as knowledge of contextual conditions.

In the end, the criterion that must be used to sort out the best roles for researchers and teachers is whether the work done by every participant contributes to the improvement of products and, in turn, the better achievement of learning goals by students. Because little precedent exists for systems in which all participants are equally committed and equally accountable, it is impossible to say exactly what roles will be optimal. Those suggested above might be a good starting point.

Conclusion

In simplest terms, our argument has been that classroom teaching will improve, for all students, only if the variation in instructional practices across classrooms is reduced and the quality of instruction improves over time. We claim that a promising approach, and one not yet tried seriously in the United States, is to jointly build instructional products that can be used by all teachers, continually tested, and refined. Although not an entirely foreign concept in American education, building public, changeable instructional products is likely to take hold only if the minimal set of features we identified here becomes part of the education culture.

NOTES

Preparation of this article was supported, in part, by the National Science Foundation, Grant 0083429 to the Mid-Atlantic Center for Teaching and Learning Mathematics. The opinions expressed in the article are those of the authors and not necessarily those of the foundation. We appreciate feedback on previous drafts of this article from the editors, several anonymous reviewers, Ron Gallimore, Jim Stigler, and the faculty and doctoral students in the mathematics education research seminar at the University of Delaware.

[1]For definitions and interpretations of these worlds of knowledge for students' classroom activities, see Bereiter (2002) and Bereiter and Scardamalia (1996).

[2]We induced the three features by looking across a number of examples both inside and outside education. Outside education, in addition to the health care example reviewed in this article, the examples include such diverse systems as the wind turbine industry in Denmark (Douthwaite, 2002) and the Xerox Corporation's Eureka project, which

improves the repair procedures for the company's machines (Bobrow & Whalen, 2002). See Morris and Hiebert (2009) for a review of these examples. Inside education, additional examples include the improvement of teacher preparation courses in mathematics at the University of Delaware (Hiebert & Morris, 2009), the Kamehameha Elementary Education Project (KEEP) in Hawaii (Tharp & Gallimore, 1988; Tharp et al., 1984), and the testing and revising of a commercial mathematics curriculum by the teachers at Public School 112M in East Harlem (El Barrio–Hunter College PDS Partnership Writing Collective, 2009).

REFERENCES

Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6–8.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum.

Bereiter, C., & Scardamalia, M. (1996). Rethinking learning. In D. R. Olson & N. Torrance (Eds.), *Handbook of education and human development: New models of learning, teaching, and schooling* (pp. 485–513). Cambridge, MA: Blackwell.

Bobrow, D. G., & Whalen, J. (2002). Community knowledge sharing in practice: The Eureka story. *Reflections*, *4*(2), 47–59.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences*, *2*(2), 141–178.

Brown, A. L. (1994). The advancement of learning. *Educational Researcher*, *23*(8), 4–12.

Bulterman-Bos, J. A. (2008). Will a clinical approach make education research more relevant for practice? *Educational Researcher*, *37*, 412–420.

Clandinin, D. J., & Connelly, F. M. (1991). Narrative and story in practice and research. In D. A. Schön (Ed.), *The reflective turn: Case studies in and on educational practice* (pp. 258–281). New York: Teachers College Press.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *31*(1), 9–13.

Cochran-Smith, M., & Lytle, S. L. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher*, *19*(2), 2–11.

Cochran-Smith, M., & Lytle, S. L. (Eds.). (1993). *Inside/outside: Teacher research and knowledge*. New York: Teachers College Press.

Cochran-Smith, M., & Lytle, S. L. (1999). The teacher researcher movement: A decade later. *Educational Researcher*, *28*(7), 15–25.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, *13*(1), 15–42.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform* (Consortium for Policy Research in Education Research Report No. RR-63). New York: Teachers College–Columbia University, Center for Continuous Instructional Improvement.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890–1990* (2nd ed.). New York: Teachers College Press.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*(3), 3–14.

Donovan, M. S., Wigdor, A. K., & Snow, C. E. (Eds.). (2003). *Strategic education research partnership*. Washington, DC: National Academy Press.

Douthwaite, B. (2002). *Enabling innovation: A practical guide to understanding and fostering technological change*. New York: Zed Books.

Eisner, E. W. (1995). Preparing teachers for schools of the 21st century. *Peabody Journal of Education*, *70*(3), 99–111.

El Barrio–Hunter College PDS Partnership Writing Collective (2009). On the unique relationship between teacher research and commercial mathematics curriculum development. In J. T. Remillard, B. A. Herbel-Eisenmann, & G. M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 118–133). New York: Routledge.

Fenstermacher, G. D. (1988). The place of science and epistemology in Schön's conception of reflective practice. In P. P. Grimmett & G. L. Erickson (Eds.), *Reflection in teacher education* (pp. 39–46). New York: Teachers College Press.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, *107*, 186–213.

Fey, J. T. (1979). Mathematics teaching today: Perspectives from three national surveys. *Arithmetic Teacher*, *27*(2), 10–14.

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher*, *18*(7), 4–10.

Gravemeijer, K., & van Eerde, D. (2009). Design research as a means for building a knowledge base for teachers and teaching in mathematics education. *Elementary School Journal*, *109*, 510–524.

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher*, *31*(5), 3–15.

Hiebert, J., & Morris, A. K. (2009). Building a knowledge base for teacher education: An experience in K–8 mathematics teacher preparation. *Elementary School Journal*, *109,* 475–490.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., et al. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 Video Study. *Educational Evaluation and Policy Analysis*, *27,* 111–132.

Hoetker, J., & Ahlbrand, W. P., Jr. (1969). The persistence of the recitation. *American Educational Research Journal*, *6,* 145–167.

Institute for Healthcare Improvement. (2007). *Measures*. Retrieved from http://www.ihi.org/IHI/Topics/Improvement/ImprovementMethods/Measures

Jacob, E., & White, C. S. (Eds.). (2002). Scientific research in education [Theme issue]. *Educational Researcher*, *31*(8).

Kelly, A. (2003). Research as design. *Educational Researcher*, *32*(1), 3–4.

Kennedy, M. (1999). Ed schools and the problem of knowledge. In J. D. Raths & A. C. McAninch (Eds.), *Advances in teacher education: Vol. 5. What counts as knowledge in teacher education?* (pp. 29–45). Stamford, CT: Ablex.

Kenney, C. (2008). *The best practice: How the new quality movement is transforming medicine*. New York: Public Affairs.

Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York: Crown.

Kristof, N. D. (2009, February 15). Our greatest national shame. *New York Times*, p. WK11.

Lamberg, T. D., & Middleton, J. A. (2009). Design research perspectives on transitioning from individual microgenetic interviews to a whole-class teaching experiment. *Educational Researcher*, *38*(4), 233–245.

Langley, G. J., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (1996). *The improvement guide: A practical approach to enhancing organizational performance*. San Francisco: Jossey-Bass.

Leonhardt, D. (2009, November 8). Dr. James will make it better. *New York Times Magazine*, 31–37, 44–46.

Lewin, K., & Grabbe, P. (1945). Conduct, knowledge, and acceptance of new values. *Journal of Social Values*, *3*, 53–64.

Lewis, C. (2002). *Lesson study: A handbook of teacher-led instructional change*. Philadelphia: Research for Better Schools.

Lewis, C., & Tsuchida, I. (1997). Planned educational change in Japan: The shift to student-centered elementary science. *Journal of Educational Policy*, *12,* 313–331.

Morris, A. K., & Hiebert, J. (2009). Introduction: Building knowledge bases and improving systems of practice. *Elementary School Journal*, *109*, 429–441.

Morris, A. K., Hiebert, J., & Spitzer, S. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can pre-service teachers learn? *Journal for Research in Mathematics Education*, *45*, 491–521.

Moss, P. A., Phillips, D. C., Erickson, F. D., Floden, R. E., Lather, P. A., & Schneider, B. L. (2009). Learning from our differences: A dialogue across perspectives on quality in education research. *Educational Researcher*, *38*, 501–517.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science*, *315*, 1795–1796.

Popper, K. (1972). *Objective knowledge*. London: Oxford University Press.

Popper, K. (1985). Piecemeal social engineering. In D. Miller (Ed.), *Popper selections* (pp. 304–318). Princeton, NJ: Princeton University Press. (Original work published 1944.)

Raudenbush, S. W. (2009). The *Brown* legacy and the O'Conner challenge: Transforming schools in the images of children's potential. *Educational Researcher*, *38*, 169–180.

Remillard, J. T., & Bryans, M. B. (2004). Teachers' orientation toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics Education*, *35*, 352–388.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Shimahara, N. K., & Sakai, A. (1995). *Learning to teach in two cultures: Japan and the United States*. New York: Garland.

Snow, C. E. (2001). Knowing what we know: Children, teachers, researchers. *Educational Researcher*, *30*(7), 3–9.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science*, *328*, 450–452.

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Stigler, J. W., & Thompson, B. J. (2009). Thoughts on creating, accumulating, and utilizing shareable knowledge to improve teaching. *Elementary School Journal*, *109*, 442–457.

Strategic Education Research Partnership. (n.d.). Algebra I. Retrieved from http://www.serpinstitute.org/about/field-sites/msan-algebra.php

Strategic Education Research Partnership. (n.d.). Internal coherence partnership and protocol. Retrieved from http://www.serpinstitute.org/about/field-sites/boston-ic-ccdd.php

Strategic Education Research Partnership. (n.d.). An introduction to Word Generation. Retrieved from http://www.serpinstitute.org/tools-and-resources/word-generation.php

Strategic Education Research Partnership. (n.d.). The SERP field site design team. Retrieved from http://www.serpinstitute.org/about/field-sites/structure-design-team

Strategic Education Research Partnership. (n.d.). The SERP field site structure. Retrieved from http://www.serpinstitute.org/about/field-sites/structure.php

Strategic Education Research Partnership. (n.d.) Word Generation: Teacher guide. Retrieved from www.serpinstitute.org/downloads/344wgoct1eacher_guide11.doc

Stuhlman, M. W., & Pianta, R. C. (2009). Profiles of educational quality in first grade. *Elementary School Journal*, *109*, 323–342.

Tharp, R. G., & Gallimore, R. (1979). The ecology of program research and evaluation: A model of evaluation succession. In L. Sechrest, S. G. West, M. A. Phillips, R. Redner, & W. Yeaton (Eds.), *Evaluation studies: Review annual* (Vol. 4, pp. 39–60). Beverly Hills, CA: Sage.

Tharp, R. G., & Gallimore, R. (1982). Inquiry processes in program development. *Journal of Community Psychology*, *10*, 103–118.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, UK: Cambridge University Press.

Tharp, R. G., Jordan, C., Speidel, G. E., Au, A. H., Klein, T. W., Calkins, R. P., et al. (1984). Product and process in applied developmental research: Education and the children of a minority. In M. E. Lamb, A. L. Brown, & R. B. Rogoff (Eds.), *Advances in developmental psychology* (*Vol. 3*). Hillsdale, NJ: Lawrence Erlbaum.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*, 83–91.

Viadero, D. (2009, May 13). Partnership's first product aimed at middle school vocabulary. *Education Week*, *28*(31), 1. Retrieved from http://www.edweek.org/

Wang-Iverson, P., & Yoshida, M. (Eds.). (2005). *Building our understanding of lesson study*. Philadelphia: Research for Better Schools.

White, C. W., & Kim, J. (2009, May). *Putting the pieces of the puzzle together: How systematic vocabulary instruction and expanded learning time can address the literacy gap*. Retrieved from http://www.americanprogress.org/issues/2009/05/elt_language_development.html

Wiliam, D. (2009). *Assessment for learning: Why, what, and how*. London: Institute of Education.

Word Generation. (n.d.) Exchange ideas. Retrieved from http://wordgeneration.org/exchange1.html

AUTHORS

ANNE K. MORRIS is a professor at the University of Delaware, School of Education, 16 W. Main Street, Newark, DE 19716; *abmorris@udel.edu.* Her research focuses on mathematics learning and teacher education.

JAMES HIEBERT is the Robert J. Barkley Professor at the University of Delaware, School of Education, 16 W. Main Street, Newark, DE 19716; *hiebert@udel.edu*. His research focuses on teaching and learning in mathematics classrooms.

Manuscript received April 22, 2010
Revisions received August 10, 2010, and October 28, 2010
Accepted November 13, 2010